論　文

非正規性の下での最尤法

渡 邊 美 智 子
山 口 和 範

0. 序

多くの統計モデルにおいて誤差項は正規分布に従う確率変数と仮定され，その下で目的母数の最尤推定が施行される。この場合，結果の理論的妥当性はデータが正規性の仮定をみたす場合のみ保障されるため，実データに適用する場合，その頑健性が一般に問題になる。この問題を回避する一つの手段として，とくに異常値の混入等が想定される場合などに，初めから誤差項に対して正規分布よりも裾の重い対称な分布を仮定し，その下で最尤推定を行う方向が考えられる。但し，解析的な煩雑さが予想されるが，EM アルゴリズムの適用により比較的容易に最尤法を施行できる。本稿の目的は，このための誤差項に与える分布のクラスを導入し，線形回帰モデル及び因子分析モデルへの応用を考察することである。

1. Normal/Independent Distributions

Andrews, et al (1972), Andrews & Mallows (1974) は，正規分布以外の誤差分布のモデルとして，裾が正規分布よりも比較的重く，単峰性かつ対称分布の族を Normal/Independent Distributions，略して，N/I 分布として次の定義を与えている：ここに，u を標準正規分布に従う確率変数，q を u と独立でかつ確率（密度）関数 $M(q)$ に従う確率変数とするとき，$Z=uq^{-1/2}$ で

定義される scaled normal な確率変数 Z は，N/I 分布をもつ。

N/I 分布の族は，$M(q)$ を具体的に与えることで，とくに次の5つの分布を含むことができる。

1. q が $(\lambda_1, \cdots \lambda_j \cdots, \lambda_m)$ をとる離散分布するとき，Z は，平均は0，分散は $1/\lambda_j$ で，分散のみ各々異なる m 個の正規分布の混合分布に従う。とくに，$m=2$, $\lambda_1=1$ のとき，contaminated normal と呼ばれる。
2. mq が自由度 m のx^2 分布に従うとき，Z は自由度 m の t 分布に従う。
3. q が自由度1の x^2 分布に従うとき，Z は Cauchy 分布に従う。
4. $M(q)=(1/2)\,q^{-2}exp.\,[-1/2q]$ とするとき，Z の分布は double exponential 分布である。
5. $M(q)=\sum_{k=1}^{\infty}(-1)^{k-1}k^2q^{-2}exp.\,[-k/2q]$ のとき，Z の分布は logistic分布である。

とくに，u を正規分布と限定しない場合でも，このような scale mixture の分布はもとの変量の分布よりも裾が重くなることが Beale & Mallows (1959)により証明されている。

2.　N/I 分布の下での線形回帰分析法

線形回帰モデル：$\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{e}$ において，$\sigma^{-1}\boldsymbol{q}$ は N/I 分布をすると仮定する。即ち，密度関数 $M(q)$ に従う互いに独立な n 個の確率変数 q_i が存在して，q_i の条件付きの下で，$\sigma^{-1}\boldsymbol{e}$ は，$N(\boldsymbol{0},\ \boldsymbol{W})$ に従う。ここに，$\boldsymbol{W}=\mathrm{diag.}\{q_1^{-1}, \cdots, q_n^{-1}\}$ である。

$\boldsymbol{q}=(q_1, \cdots, q_n)$ の条件付きの下で，モデルのパラメータ $\boldsymbol{\beta}$ は，通常の重み付き最小自乗法により推定される。ここでは，$\boldsymbol{q}$ は観測不可能な数量，即ち，欠測値として取り扱う。Dempster, Laird & Rubin (1977) による EM アルゴリズムの記法によれば，$\boldsymbol{O}=(\boldsymbol{q},\ \boldsymbol{Y})$ が観測不可能な完全データ，$\boldsymbol{Y}$ が観測可能な不完全データとなる。$M(q)$ が，$\boldsymbol{\beta}, \sigma$ とは無関係であるという仮定の下で，観測データ $\boldsymbol{Y}$ に基づく $\boldsymbol{\beta}, \sigma$ の最尤推定値の導出がここでの目的であ

る。これは，以下に与える $\boldsymbol{O}$ の対数尤度の $\boldsymbol{Y}$ が与えられた下での条件付き期待値の最大化を通して実現される。

$$L(\boldsymbol{\beta}, \sigma)=\log f(\boldsymbol{O})$$
$$=(-n/2)\log(2\pi)-n\log\sigma+(1/2)\sum_{i=1}^{n}\log(q_i)-(1/2\sigma^2)\sum_{i=1}^{n}q_i(Y_i-\boldsymbol{X}_i\boldsymbol{\beta})^2$$
$$+\sum_{i=1}^{n}\log M(q_i)$$

$L(\boldsymbol{\beta}, \sigma)$ の条件付期待値の評価を E-step において，また，目的パラメータに関する最大化を M-step において，各々実現する。

E-step：観測データ $\boldsymbol{Y}$，パラメータの暫定値 $(\boldsymbol{\beta}^{(k)}, \sigma^{(k)})$ の下で，$L(\boldsymbol{\beta},\sigma)$ の条件付期待値 $E[L(\boldsymbol{\beta}, \sigma) \mid \boldsymbol{Y}, (\boldsymbol{\beta}^{(k)}, \sigma^{(k)})]$ を評価する。即ち，

$$q_i^{(k+1)}=E[q_i \mid \boldsymbol{Y}, (\boldsymbol{\beta}^{(k)}, \sigma^{(k)})]$$

を求めることに他ならない。

M-step：E-step で求めた $q_i^{(k+1)}$ が与えられたものとして，$L(\boldsymbol{\beta}, \sigma)$ のパラメータに関する最大化を行う。即ち，

$$\boldsymbol{\beta}^{(k+1)}=(\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{X})^{-1}\ \boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{Y},$$
$$\sigma^{(k+1)}=(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}^{(k+1)})'\ \boldsymbol{W}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}^{(k+1)})/n$$

により，パラメータの推定値を更新する。

上記の E-step，M-step を適当な初期値の下に反復実行し，目的パラメータの最大推定値を得る。E-step は，q に関する確率（密度）関数 $M(q)$ を特定することにより，具体化される。以下にその 2 つの例を挙げる。

EX 1.　(Contaminated Normal Case)

$$M(q)=\begin{cases}1-\delta & , \text{ if } q=1, \\ \delta & , \text{ if } q=\lambda, \\ 0 & , \text{ otherwise,}\end{cases}$$
ここに，$\lambda<<1$，$0<\delta<1$，λ, δ は既知。

$M(q)$ を上記の様に特定すれば，$\sigma^{-1}e_i$ の分布として，$N(0,1)$ と $N(0, 1/\lambda)$ の混合分布を仮定したこととなる。

ここで，$\boldsymbol{Y}$ とパラメータの暫定値$(\boldsymbol{\beta}^{(k)}, \sigma^{(k)})$ が与えられた下での q の条件付分布を具体的に評価して，

$$q_i^{(k+1)}=E[q_i \mid \boldsymbol{Y},\ (\boldsymbol{\beta}^{(k)},\ \sigma^{(k)})]$$
$$=[1-\delta+\delta\lambda^{1+1/2}exp\{(1-\lambda)Z_i^2/2]/[1-\delta+\delta\lambda^{1/2}exp.\{(1-\lambda)\ Z_i^2/2]$$

をえる。ここに，$Z_i=(y_i-\boldsymbol{X}_i\boldsymbol{\beta}^{(k)})/\sigma^{(k)}$ で与えられる。

EX 2.　(Univariate t Case)

qm を自由度 m の x^2 分布とおくと，$\sigma^{-1}e_i$ は，自由度 m の t 分布に従うことになる。このとき，

$$q_i^{(k+1)}=E[q_i \mid \boldsymbol{Y},\ (\boldsymbol{\beta}^{(k)},\ \sigma^{(k)})]$$
$$=(m+1)/(m+Z_i^2)$$

をえる。ここに，$Z_i=(y_i-\boldsymbol{X}_i\boldsymbol{\beta}^{(k)})/\sigma^{(k)}$，である。

上記の $q_i^{(k+1)}$ は，何れも，ロバスト回帰分析で用いられる重み関数 $w(Z)$ の性質：正値，単調非増加，$w(Z_i)=w(-Z_i)$ をみたしている。ロバスト回帰分析では，適当な重み関数を与えて，重み付き最小自乗法と残差 Z の計算，その下での重みの更新を繰り返すアルゴリズム：IRLS (Iteratively Reweighted Least Squares) が頻用されている。

一般に，$Z=\sigma^{-1}e$ に適当な分布 $d(Z)$ を仮定した際に，IRLS による推定量が最尤推定量となるための条件は，重み関数が以下で与えられることが Andrews(1974), Beaton *et. al.*(1974) によって導かれている：

$$w(Z)=\begin{cases} -d'(Z)/Zd(Z) & \text{for } Z\neq 0, \\ \lim\limits_{Z\to 0} -d'(Z)/Zd(Z) & \text{for } Z=0. \end{cases}$$

従って，先述の $q_i^{(k+1)}$ は上式から直接求めることもできる。

Dempster, Laird & Rubin(1980) において，とくに $d(Z)$ が N/I 分布の場合に一般に成立する性質として，以下が与えられている：

$Z\neq 0$ に対して，

(1)　Z の下での q の条件付分布は存在する。

(2)　$E[q^k \mid Z]<\infty$, for $k>-1/2$

(3)　$w(Z)=E[q \mid Z]$

(4)　$w'(Z)=-Z\ \mathrm{var}\ (q \mid Z)$

（５） $w(Z)=w(-Z)$, $w(Z)$ は，正の有界な単調非増加関数

（６） $Z=0$ に対して，Z の下での q の条件付分布は，$E[q^{1/2}]<\infty$ のとき，またそのときに限り存在する。

（７） $w(0)$ は $E[q^{3/2}]<\infty$ のとき，またそのときに限り存在し，かつ有限である。また，$Z\neq 0$ に対して，$w(0)>w(Z)$

（８） $w'(0)$ は $E[q^{5/2}]<\infty$ のとき，またそのときに限り存在しかつ有限である。

上記から，q の条件付期待値が，重み関数として有効に作用することがわかる。

3.　因子分析モデルにおけるロバスト推定法

一般に，多因子モデルは，

$$\boldsymbol{Y}=\boldsymbol{\alpha}+\boldsymbol{\beta}\boldsymbol{Z}+\boldsymbol{e}, \qquad (1)$$

で表現される。ここに，$\boldsymbol{Y}$ は $(p\times 1)$ の各個人の応答ベクトル，$\boldsymbol{Z}$ は $(q\times 1)$ の共通因子に関する因子得点ベクトル，$\boldsymbol{\alpha}$ は $(p\times 1)$ の $\boldsymbol{Z}$ の平均ベクトル，$\boldsymbol{\beta}$ は $(p\times q)$ の因子負荷行列，$\boldsymbol{e}$ は $(p\times 1)$ の特殊因子ベクトル又は，誤差ベクトルで，$\boldsymbol{Z}$ と $\boldsymbol{e}$ は独立を仮定する。

通常，因子分析モデルにおける最尤推定法は，共通因子 $\boldsymbol{Z}$ と特殊因子 $\boldsymbol{e}$ に関する多変量正規性の仮定の下で構築される。これに関して，Flettcher-Powell 型や Newton-Raphson 型等種々の方向から，最尤推定量の求解アルゴリズムが提唱されている。上述の EM アルゴリズムの観点からも，Rubin & Thayer (1982) が一つの方法を提唱している。しかしながら，彼らの論文の最後の一文：Of course, the entire issue of the sensitivity of results to the assumption of multivariate normality is important for the wise application of thetechnique in practice. にもみられるように，実データの分析に無条件に多変量正規性の仮定おくことにはやや無理があろう。

本節では，多変量正規性を前述の N/I 分布でおきかえ，比較的外れ値の影

響を受けにくい推定アルゴリズムの構築を試みる。また，提唱されるアルゴリズムにより導かれる推定値は，もしデータに外れ値が含まれず，即ち，多変量正規性の仮定が有効な下では，正規性を仮定したときの最尤推定値に近い値であることが期待できる。

多因子モデル(1)は，多変量（p 変量）重回帰モデルにおいて，説明変数の観測値がすべて欠測している場合とみなすことができる。つまり，対象集団の各個人は，本来 $\boldsymbol{Y}$, $\boldsymbol{Z}$ についての応答値をもち，その $\boldsymbol{Y}$ と $\boldsymbol{Z}$ の間の関係が，(1)の回帰モデルで説明されると考える。この場合，データに外れ値が含まれる可能性のモデルとして，次の2つが考えられる：

モデル(イ)：集団が最初から均質ではなく，一部の異常な能力を有する個体の存在が想定される，つまり，$\boldsymbol{Y}$, $\boldsymbol{Z}$ ともに，他から外れた値が混入している場合。

モデル(ロ)：集団の潜在能力自身は均質であるが，顕在応答 $\boldsymbol{Y}$ が観測される時点で，外れ値の混入がある場合。もともと，特殊因子は，誤差も含めて多くの要因の混合の結果であるから，特殊因子の分布の仮定としては，単一正規分布よりも，正規分布の scale mixture 等の N/I 分布をあてるほうがより現実的なモデルであろう。

［モデル(イ)の場合］

$\boldsymbol{Z}$ と $\boldsymbol{e}$ は各々 N/I 分布をする。つまり，ある $M(q)$ に従う正の確率変数 q が存在して，q が与えられた下で，$\boldsymbol{Z}$ は，平均 $\mathbf{0}$，共分散行列 $\boldsymbol{R}/q$ の正規分布に従い，また，$\boldsymbol{e}$ は平均 $\mathbf{0}$，共分散行列 $\boldsymbol{\Psi}/q$，$\boldsymbol{\Psi}=\mathrm{diag}\ \{\psi_1, \cdots, \psi_p\}$ の正規分布に従う。q の下で，$\boldsymbol{W}'=(\boldsymbol{Y}', \boldsymbol{Z}')$ は，平均 $(\boldsymbol{\alpha}', \mathbf{0})'$，共分散行列 $\sum/q$ の正規分布となる。ここに，

$$\Sigma=\begin{pmatrix}\Sigma_{YY} & \Sigma_{YZ}\\ \Sigma_{ZY} & \Sigma_{ZZ}\end{pmatrix}\equiv\begin{pmatrix}\boldsymbol{\beta R}'+\boldsymbol{\Psi} & \boldsymbol{\beta R}\\ \boldsymbol{R\beta}' & \boldsymbol{R}\end{pmatrix}$$

である。

EM アルゴリズムにおける完全データは，$\boldsymbol{X}=(q,\ \boldsymbol{Y},\ \boldsymbol{Z})$ で，不完全データは $\boldsymbol{Y}$ を指す。この場合，完全データにおける十分統計量は

$$q,\ q\boldsymbol{Y},\ q\boldsymbol{Z},\ \ q\boldsymbol{YY}',\ \ q\boldsymbol{ZY}'\ \ q\boldsymbol{ZZ}' \tag{2}$$

である。

この場合の E-step，M-step を各々，以下で具体化する：

<u>E-step</u>：$\boldsymbol{Y}$ と $\boldsymbol{\theta}^{(k)}=\{\boldsymbol{\alpha}^{(k)},\ \boldsymbol{\beta}^{(k)},\ \boldsymbol{R}^{(k)},\ \boldsymbol{\Psi}^{(k)}\}$ のもとで，(2)の各々の条件付期待値を求める。

$$E(q\,|\,\boldsymbol{Y};\ \boldsymbol{\theta}^{(k)})\equiv w^{(k)}$$

$$\begin{aligned}E(q\boldsymbol{Z}\,|\,\boldsymbol{Y};\boldsymbol{\theta}^{(k)})&=E(qE(\boldsymbol{Z}|q,\boldsymbol{Y};\boldsymbol{\theta}^{(k)})\,|\,\boldsymbol{Y};\ \boldsymbol{\theta}^{(k)})\\&=w^{(k)}E(\boldsymbol{Z}|q,\boldsymbol{Y};\boldsymbol{\theta}^{(k)})\equiv w^{(k)}\hat{\boldsymbol{Z}}^{(k)}\end{aligned} \tag{3}$$

$$\begin{aligned}E(q\boldsymbol{ZZ}'\,|\,\boldsymbol{Y};\boldsymbol{\theta}^{(k)})&=E(qE(\boldsymbol{ZZ}'|q,\ \boldsymbol{Y};\ \boldsymbol{\theta}^{(k)})\,|\,\boldsymbol{Y};\boldsymbol{\theta}^{(k)})\\&=E(q\{\hat{\boldsymbol{Z}}^{(k)}\hat{\boldsymbol{Z}}^{(k)\prime}+Cov(\boldsymbol{Z}|q;\ \boldsymbol{Y},\ \boldsymbol{\theta}^{(k)})\}\,|\,\boldsymbol{Y};\boldsymbol{\theta}^{(k)})\\&=w^{(k)}\hat{\boldsymbol{Z}}^{(k)}\hat{\boldsymbol{Z}}^{(k)\prime}+\Sigma_{ZZ}^{*(k)},\end{aligned}$$

ここに，

$$\hat{\boldsymbol{Z}}^{(k)}=\Sigma_{ZY}^{(k)}\Sigma_{YY}^{(k)-1}(\boldsymbol{Y}-\boldsymbol{\alpha}^{(k)}),$$

$$\Sigma_{ZZ}^{*(k)}=\Sigma_{ZZ}^{(k)}-\Sigma_{YZ}^{(k)}\Sigma_{YY}^{(k)-1}\Sigma_{YZ}^{(k)}\ \text{である。}$$

具体的に確率（密度）関数 $M(q)$ を与えて，$w^{(k)}$を評価すればよい。

EX 1.　(Contaminated Multivaviate Normal Case)

$$M(q)=\begin{cases}1-\delta & \text{if } q=1\\ \delta & \text{if } =\lambda\\ 0 & \text{otherwise, ここに，} \lambda\ll 1.\end{cases}$$

$M(q)$ を上記で特定すると，

$\begin{pmatrix}\boldsymbol{Y}\\ \boldsymbol{Z}\end{pmatrix}$の従う分布は，$N\left(\begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{0}\end{pmatrix},\ \Sigma\right)$ と $N\left(\begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{0}\end{pmatrix},\ \Sigma/\lambda\right)$ の混合分布となる。

この場合，

$w=[1-\delta+\delta\lambda^{1+p/2}exp\}(1-\lambda)d^2/2\}]/[1-\delta+\delta\lambda^{p/2}exp\{(1-\lambda)d^2/2\}]$,

となる。ここに，

$d^2=(\boldsymbol{Y}-\boldsymbol{\alpha})'\sum_{YY}^{-1}(\boldsymbol{Y}-\boldsymbol{\alpha})$ である。

EX 2.　(Multivaviate t Case)

$mq \sim x_m^2$ のとき，

$\begin{pmatrix}\boldsymbol{Y}\\ \boldsymbol{Z}\end{pmatrix}$ は $t\ \left(\begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{0}\end{pmatrix},\ \sum,\ m\right)$，つまり，平均ベクトル $\begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{0}\end{pmatrix}$，分散共分散行列$\sum$，自由度 m の多変量 t 分布に従う。ここに，$\boldsymbol{x} \sim t(\boldsymbol{\mu}, \sum, m)$，とは，Dunnett & Sobel (1954, 1955) の定義によれば，$\boldsymbol{x}$ が $\boldsymbol{\mu}+s^{-1}\boldsymbol{y}$ と同じ分布をもつことである。ここに，$\boldsymbol{y}$ と s^2 は独立，$\boldsymbol{y} \sim N(\boldsymbol{0}, \sum)$，$ms^2 \sim x^2{}_m$ である。

この場合，$w=(m+p)/(m+d^2)$ となる。

<u>*M*-step</u>：*E*-step で評価した(3)の値を用いて，

$$\log L(\boldsymbol{\theta})=\log\prod_{i=1}^{n} f(q_i, \boldsymbol{Y}_i, \boldsymbol{Z}_i)$$

$$=C-\frac{n}{2}\log|\boldsymbol{\Psi}|\cdot|\boldsymbol{R}|+\frac{p}{2}\sum_{i=1}^{n}\log q_{\ i}-\frac{1}{2}\sum_{i=1}^{n}tr.\ [q_i\{\boldsymbol{\Psi}^{-1}($$

$$(\boldsymbol{Y}_i-\boldsymbol{\alpha})(\boldsymbol{Y}_i-\boldsymbol{\alpha})'-(\boldsymbol{Y}_i-\boldsymbol{\alpha})\boldsymbol{Z}_i'\boldsymbol{\beta}'+\boldsymbol{\beta}\boldsymbol{Z}_i\boldsymbol{Z}_i'\boldsymbol{\beta}')+\boldsymbol{R}^{-1}\boldsymbol{Z}_i\boldsymbol{Z}_i'\}]$$

の条件付期待値の最大化をパラメータ間の種々の制約の下に行う。

因子分析モデルでは，モデル自身の identifiability を保障するために，パラメータ間に制約が付されるのが普通である。ここでは制約式を Rubin & Thayer (1982) より一般化した下で，統一的な推定法を導く：

パラメータの制約として，

（i）$\boldsymbol{\alpha}$ が部分的に既知

（ii）$\boldsymbol{\beta}$　　〃

（iii）$\boldsymbol{R}$ は既知または未知

（iv）　$\{\psi_1, \psi_2, \cdots, \psi_p\}$ は，同じ値をもついくつかの群に分かれ，そのうちのある群は既知の値をもつ。

の4通りの組み合せを考える。

予め，以下の記法を定義しておく：

$$\begin{pmatrix} \boldsymbol{S}_{YY} & \boldsymbol{S}_{YZ} \\ \boldsymbol{S}_{ZY} & \boldsymbol{S}_{ZZ} \end{pmatrix} \fallingdotseq \begin{pmatrix} \sum_{i=1}^{n} w_i \boldsymbol{Y}_i \boldsymbol{Y}_i'/w_0 & \sum_{i=1}^{n} w_i \boldsymbol{Y}_i \boldsymbol{Z}_i'/w_0 \\ \sum_{i=1}^{n} w_i \boldsymbol{Z}_i \boldsymbol{Y}_i'/w_0 & \sum_{i=1}^{n} w_i \boldsymbol{Z}_i \boldsymbol{Z}_i'/w_0 \end{pmatrix},$$

$$\begin{pmatrix} \boldsymbol{C}_{YY} & \boldsymbol{C}_{YZ} \\ \boldsymbol{C}_{ZY} & \boldsymbol{C}_{ZZ} \end{pmatrix} \fallingdotseq \begin{pmatrix} \boldsymbol{S}_{YY} - \overline{\boldsymbol{Y}}\overline{\boldsymbol{Y}}' & \boldsymbol{S}_{YZ} - \overline{\boldsymbol{Y}}\overline{\boldsymbol{Y}}' \\ \boldsymbol{S}_{ZY} - \overline{\boldsymbol{Z}}\overline{\boldsymbol{Y}}' & \boldsymbol{S}_{ZZ} - \overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}' \end{pmatrix},$$

ここに，$\overline{\boldsymbol{Y}} = \sum_{i=1}^{n} w_i \boldsymbol{Y}_i/w_0$，$\overline{\boldsymbol{Z}} = \sum_{i=1}^{n} w_i \boldsymbol{Z}_i/w_0$，$w_0 = \sum_{i=1}^{n} w_i.$

（1）　$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}^{(1)} \\ \boldsymbol{\alpha}^{(2)} \end{pmatrix}$，$\boldsymbol{\alpha}^{(1)}$ は未知，$\boldsymbol{\alpha}^{(2)}$ は既知，$\boldsymbol{\beta}$ は全て未知のときの場合。

$$\hat{\boldsymbol{R}} = \boldsymbol{S}_{ZZ} w_0/n,\ \hat{\boldsymbol{\beta}}^{(1)} = \boldsymbol{C}_{Y^{(2)}Z} \boldsymbol{C}^{-1}{}_{ZZ},\ \hat{\boldsymbol{\beta}}^{(2)} = \boldsymbol{C}^*{}_{Y^{(2)}ZZ} \boldsymbol{S}^{-1}{}_{ZZ},$$

$$\hat{\boldsymbol{\alpha}}^{(1)} = \overline{\boldsymbol{Y}}^{(1)} - \boldsymbol{\beta}^{(1)} \boldsymbol{Z},$$

$$\hat{\boldsymbol{\Psi}}^{(1)} = \mathrm{diag.}\{\boldsymbol{C}_{Y^{(1)}Y^{(1)}} - \boldsymbol{C}_{Y^{(1)}Z} \boldsymbol{C}^{-1}{}_{ZZ} \boldsymbol{C}_{ZY^{(1)}}\}/n,$$

$$\hat{\boldsymbol{\Psi}}^{(2)} = \mathrm{diag.}\{\boldsymbol{C}^*{}_{Y^{(2)}Y^{(2)}} - \boldsymbol{C}^*{}_{Y^{(2)}Z} \boldsymbol{S}^{-1}{}_{ZZ} \boldsymbol{C}^*{}_{ZY^{(2)}}\}/n,$$

ここに　$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}^{(1)} \\ \boldsymbol{Y}^{(2)} \end{pmatrix}$，$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \end{pmatrix}$，$\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}^{(1)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Psi}^{(2)} \end{pmatrix}$ は，$\begin{pmatrix} \boldsymbol{\alpha}^{(1)} \\ \boldsymbol{\alpha}^{(2)} \end{pmatrix}$ に対応する分割、

$$\boldsymbol{C}^*{}_{Y^{(2)}Y^{(2)}} = \boldsymbol{C}_{Y^{(2)}Y^{(2)}} + (\overline{\boldsymbol{Y}}^{(2)} - \boldsymbol{\alpha}^{(2)})(\overline{\boldsymbol{Y}}^{(2)} - \boldsymbol{\alpha}^{(2)})',$$

$$\boldsymbol{C}^*{}_{Y^{(2)}Z} = \boldsymbol{C}_{Y^{(2)}Z} + (\overline{\boldsymbol{Y}}_{(2)} - \boldsymbol{\alpha}_{(2)})'\overline{\boldsymbol{Z}},$$

である。

$\boldsymbol{\Psi}$ は，各要素を推定し，予め規定した同一群ごとにその平均値を求めればよい。

（2）

$$\boldsymbol{\alpha}=\begin{pmatrix}\boldsymbol{\alpha}^{(1)}\\ \boldsymbol{\alpha}^{(2)}\end{pmatrix},\ \boldsymbol{\alpha}^{(2)}：既知，\ \boldsymbol{\beta}=\begin{pmatrix}\boldsymbol{\beta}^{(1)}\\ \boldsymbol{\beta}^{(2)}\end{pmatrix}=\begin{pmatrix}\vdots\\ \boldsymbol{\beta}^{(1)}{}_{\nu}\\ \vdots\\ \boldsymbol{\beta}^{(2)}{}_{\rho}\\ \vdots\end{pmatrix}$$ の場合ここに，

$\boldsymbol{\beta}^{(1)}{}_{\nu}{}^{(1)}$：未知

$\boldsymbol{\beta}^{(1)}{}_{\nu}{}^{(2)}$：既知

$\boldsymbol{\beta}^{(2)}{}_{\rho}{}^{(1)}$：未知

$\boldsymbol{\beta}^{(2)}{}_{\rho}{}^{(2)}$：既知とする。

$$\hat{\boldsymbol{R}}=\boldsymbol{S}_{ZZ},$$

$$\hat{\boldsymbol{\beta}}^{(1)}{}_{\nu}{}^{(1)}=(\boldsymbol{C}_{Y^{(1)}{}_{\nu}Z^{(1)}}-\boldsymbol{\beta}^{(1)}{}_{\nu}{}^{(2)}\boldsymbol{C}_{Z^{(2)}Z^{(1)}})\boldsymbol{C}^{-1}{}_{Z^{(1)}Z^{(2)}},$$

$$\hat{\boldsymbol{\beta}}^{(2)}{}_{\rho}{}^{(1)}=(\boldsymbol{C}^{*}{}_{Y^{(2)}{}_{\rho}Z^{(1)}}-\boldsymbol{\beta}^{(2)}{}_{\rho}{}^{(2)}\boldsymbol{S}_{Z^{(2)}Z^{(1)}})\boldsymbol{S}^{-1}{}_{Z^{(1)}Z^{(2)}},$$

$$\hat{\boldsymbol{\alpha}}^{(1)}=\overline{\boldsymbol{Y}}-\boldsymbol{\beta}^{(1)}\overline{\boldsymbol{Z}},$$

$$\hat{\boldsymbol{\Psi}}^{(1)}=\mathrm{diag.}(\boldsymbol{C}_{Y^{(1)}Y^{(2)}}-2\boldsymbol{C}_{Y^{(1)}Z}\boldsymbol{\beta}'^{(1)}+\boldsymbol{\beta}^{(1)}\boldsymbol{C}_{ZZ}\boldsymbol{\beta}'^{(1)})$$

$$\hat{\boldsymbol{\Psi}}^{(2)}=\mathrm{diag.}(\boldsymbol{C}^{*}{}_{Y^{(2)}Y^{(2)}}-2\boldsymbol{C}^{*}{}_{Y^{(1)}Z}\boldsymbol{\beta}'^{(2)}+\boldsymbol{\beta}^{(2)}\boldsymbol{S}_{ZZ}\boldsymbol{\beta}'^{(2)})$$

ここに，$\boldsymbol{Z}=\begin{pmatrix}\boldsymbol{Z}^{(1)}\\ \boldsymbol{Z}^{(2)}\end{pmatrix}$ は $\boldsymbol{\beta}_{\nu\rho}{}^{(1)}$，$\boldsymbol{\beta}_{\nu\rho}{}^{(2)}$ に対応する分割である。

［モデル(ロ)の場合］

多因子モデル(1)に対して，$Z\sim N(\boldsymbol{0},\ \boldsymbol{R})$，そして，$\boldsymbol{e}$ のみ N/I 分布を仮定する。つまり，$q\sim M(q)$ に対して，q の条件付の下で，$\boldsymbol{e}\sim N(\boldsymbol{0},\ \boldsymbol{\Psi}/q)$ とする。

$$\Sigma=\begin{pmatrix}\Sigma_{YY} & \Sigma_{YZ}\\ \Sigma_{ZY} & \Sigma_{ZZ}\end{pmatrix}=\begin{pmatrix}\boldsymbol{\beta R\beta}'+\boldsymbol{\Psi}/q & \boldsymbol{\beta R}\\ \boldsymbol{R\beta}' & \boldsymbol{R}\end{pmatrix}$$

とおくとき，q の下で，$\boldsymbol{W}=\begin{pmatrix}\boldsymbol{Y}\\ \boldsymbol{Z}\end{pmatrix}$は，$N\left(\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{0}\end{pmatrix},\ \Sigma\right)$に従う。

このとき，完全データ　$\boldsymbol{X}=(q,\ \boldsymbol{Y},\ \boldsymbol{Z})$ に対して，十分統計量は，

$$q,\ q\boldsymbol{Y},\ q\boldsymbol{Z},\ q\boldsymbol{YY},\ q\boldsymbol{ZY}',\ q\boldsymbol{ZZ}',\ \boldsymbol{ZZ}' \qquad (4)$$

である。

<u>E-step</u>：　$\boldsymbol{Y}$, $\boldsymbol{\theta}^{(k)}=(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\Psi}^{(k)}, \boldsymbol{R}^{(k)})$ の下での(4)の条件付期待値の計算を行う。モデル(イ)の場合と異なり，$E(\boldsymbol{Z}|q, \boldsymbol{Y}; \boldsymbol{\theta}^{(k)})$ 等が，q の係数となるため，少し，複雑となる。

EX. 1. (Contaminated Normal Case)

$$M(q)=\begin{cases}1-\delta, & \text{if } q=1,\\ \delta, & \text{if } q=\lambda,\\ 0, & \text{otherwise.}\end{cases}$$

ここに，$\lambda \ll 1$.

このとき，q の条件付分布 $h(q|\boldsymbol{Y}, \boldsymbol{\theta})$ は，以下で与えられる。

$$h(1|\boldsymbol{Y}, \boldsymbol{\theta})=(1-\delta)|\textstyle\sum^*|^{-1/2}/[(1-\delta)|\sum^*|^{-1/2}+\delta|\sum|^{-1/2}\boldsymbol{D}^2]\fallingdotseq p_1$$

$$h(\lambda|\boldsymbol{Y}, \boldsymbol{\theta})=\delta|\textstyle\sum|^{-1/2}\boldsymbol{D}^2/[(1-\delta)|\sum^*|^{-1/2}+\delta|\sum|^{-1/2}\boldsymbol{D}^2]\fallingdotseq p_2$$

ここに，　$\sum=\boldsymbol{\beta R \beta}'+\boldsymbol{\Psi}/\lambda$, $\sum^*=\boldsymbol{\beta R\beta}'+\boldsymbol{\Psi}$,

$$\boldsymbol{D}^2=exp.[(1/2)\{(\lambda^{-1}\boldsymbol{\Psi}-\boldsymbol{\Psi})\cdot\textstyle\sum^{-1}\}\cdot d^2],$$

$$d^2=(\boldsymbol{y}-\boldsymbol{\alpha})\textstyle\sum^{*-1}(\boldsymbol{y}-\boldsymbol{\alpha})'$$

である。

上記の条件付確率 p_1, p_2 を利用して，以下を得る。

$$E(q|\boldsymbol{Y}, \boldsymbol{\theta})=[(1-\delta)|\textstyle\sum|^{-1/2}+\lambda\delta|\sum^*|^{-1/2}\boldsymbol{D}^2]/[(1-\delta)|\sum|^{-1/2}+\delta|\sum^*|^{-1/2}\boldsymbol{D}^2]\fallingdotseq w$$

$$E(q\boldsymbol{Z}|\boldsymbol{Y}, \boldsymbol{\theta})=\boldsymbol{\beta R}'[p_1\textstyle\sum^{*-1}+\lambda p_2\sum^{-1}](\boldsymbol{Y}-\boldsymbol{\alpha})$$

$$E(q\boldsymbol{ZZ}'|\boldsymbol{Y}, \boldsymbol{\theta})=w\boldsymbol{R}+\boldsymbol{R\beta}'[p_1\textstyle\sum^{*-1}(-\boldsymbol{I}+(\boldsymbol{Y}-\boldsymbol{\alpha})(\boldsymbol{Y}-\boldsymbol{\alpha})'\sum^{*-1}) +\lambda p_2\sum^{-1}(-\boldsymbol{I}+(\boldsymbol{Y}-\boldsymbol{\alpha})(\boldsymbol{Y}-\boldsymbol{\alpha})'\sum^{-1})]\boldsymbol{R\beta},$$

$$E(\boldsymbol{ZZ}'|\boldsymbol{Y}, \boldsymbol{\theta})=\boldsymbol{R}+\boldsymbol{R\beta}'[p_1\textstyle\sum^{*-1}(-\boldsymbol{I}+(\boldsymbol{Y}-\boldsymbol{\alpha})(\boldsymbol{Y}-\boldsymbol{\alpha})'\sum^{*-1}) +p_2\sum^{-1}(-\boldsymbol{I}+(\boldsymbol{Y}-\boldsymbol{\alpha})(\boldsymbol{Y}-\boldsymbol{\alpha})'\sum^{-1})]\boldsymbol{R\beta},$$

<u>M-step</u>：　$\hat{\boldsymbol{R}}=\sum_{i=1}^{n}E(\boldsymbol{ZZ}'|\boldsymbol{Y}, \boldsymbol{\theta})/n$, 他はモデル(イ)の場合と同じになる。

多変量 t 分布の場合は，E-step において，数値積分が必要になる。なだらかにスソをひくような分布をあてはめたいときには，分散行列の異なる複数の

正規分布の混合を考えた方が，計算上は有効であろう。

4. 結　　言

本稿において，データに外れ値等が含まれる場合の誤差項に関する分布型の仮定として正規分布よりも適当と考えられる一つの分布族を *N/I* 分布として導入した。また，その下での最尤法を一般の線形回帰モデルにおいて概説し，更に，因子分析モデルにおける適用を提唱した。

ここでの最尤法構築のための接近法は，従来，正規性の仮定に縛られていた多くの統計手法に応用が可能である。また，目的パラメーターの推定値とともに，個々のデータに関する当該モデルからの乖離度を測る数値が重み関数の最終的な実現値として得られるため，データ自身の診断にも有効である。これらの数値的な検証が今後の課題となる。

References

[1] Andrews (1974) : Some Monte Carlo results on robust-resistant regression. *In Critical Evaluation of Pysical Structural Information,* pp. 36-44. National Academy of Sciences, Washington.

[2] Andrews & Mallows (1974) : Scale Mixtures of Normal Distributions, *J.R.S.S. B* B 36 pp. 99-102.

[3] Beal & Mallows (1959): Scale Mixing of Symmetric Distributions with Zero Means, *Ann. Math. Stat.* 40 pp. 1145-1151.

[4] Campbell (1980) : Robust Procedures in Multivariate Analysis I. Robust Covariance Estimation, *Appl. Statist.* 29 pp. 231-237.

[5] Dempster, Laird & Rubin (1977) : Maximum Likelihood from Incomplete Data via the EM Algorithm, *J.R.S.S. B* 39 pp. 1-38.

[6] Dempster, Laird & Rubin (1980): Iteratively Reweighted Least Squares for Linear Regression when Errors are Normal/Independent Distributed, *Multivariate Analysis* V. pp. 35-57.

[7] Little (1988): Robust Estimation of the Mean and Covariance matrix from data with Missing Values, *Appl. Statist.* 37 pp. 23-38.

[8] Pettitt (1985) : Re-weighted Least Squares Estimation with Censored and Grouped Data : An Application of the EM Aigorithm, *J.R.S.S. B* 47 pp. 253-260.

[9] Rubin (1983): Iteratively Reweighted Least Squares, *Entry in Encyclopedia of the Statistical Sciences,* vol. 4 Wiley pp. 272-275.

[10] Rubin & Thayer (1982): EM Algorithms for ML Factor Analysis, *Psychometrika* 47 pp. 69-75.

[11] West (1984): Outlier Models and Prior Distributions in Bayesian Linear Regression, *J.R.S.S. B* 46 pp. 431-439.